

# Measuring the impact of digitized theses: a case study from the London School of Economics

*Based on a breakout session presented at the 39th UKSG Conference, Bournemouth, April 2016*

This study tests the assertion that the online dissemination of theses has a positive impact on the research profile of the institution and sets out to gain a greater understanding of how digital theses fit into the scholarly resources landscape. The year-long study combined primary and secondary research and was undertaken with the London School of Economics, based on its programme of theses digitization. The paper outlines the types of metrics an institution may use to measure the impact of its corpus of digitized dissertations and examines how these metrics may be generated. Findings included: a higher volume of theses attracts more traffic; Google's strong indexing capabilities make it the most frequently-used tool for discovery of digital theses; primary conclusions are that there is little correlation between downloads and citations of digitized theses; having a digital thesis collection enhances the reputation of the institution; although they recognize that digital theses are a valuable research tool, postgraduates and academics widely believe that making them available affects future publication opportunities; building and maintaining a digital thesis collection makes considerable 'hidden' work for librarians in terms of training about copyright and permissions. Some conclusions: better statistics are needed, especially of citations; institutions need to promote digital thesis collections better; more work needs to be done on whether digitizing theses impairs authors' chances of traditional publication and on how digital theses affect and are affected by the open access movement.

## Background

This paper was inspired by the discovery at two UK universities, the University of Surrey and the London School of Economics and Political Science (LSE), that following a pilot project by ProQuest to digitize theses at these institutions, digital theses in their respective collections scored an extraordinarily high number of downloads (when compared, say, with e-journals or e-books). The aim of the research was to gain a greater understanding of how digital theses, clearly an important academic resource, are used and how they fit into the scholarly resources landscape. The research focused particularly on the LSE collection of theses because they had been digitized most recently and were from a smaller, more interrelated group of academic disciplines than the Surrey collection. The LSE ranks 35<sup>th</sup> in the QS Global University Rankings for 2015<sup>1</sup>.

A literature search revealed that not much prior work had been carried out in this specific area,<sup>2</sup> although there are several published articles on some of the challenges associated with setting up digital thesis collections. Some of these have been cited in this paper.

Usage statistics – measuring full-text PDF downloads both from the LSE's own institutional repository and the ProQuest PQDT (ProQuest Dissertations and Theses) database – provided the main quantitative basis for the study. Where they were available, citation statistics were also used. Comparisons were made with statistics supplied by Surrey. Qualitative information to complement the statistics was obtained by carrying out three focus groups at the LSE, with undergraduates, postgraduates and librarians respectively, and by means of four semi-structured telephone calls with LSE academics working in different disciplines.



LINDA BENNETT

Independent  
Researcher  
Gold Leaf



DIMITY FLANAGAN

Academic Support  
Librarian  
LSE

## Growth in content correlated to growth in usage

The LSE digitization project commenced in 2014. By May 2015, 2,000 digitized theses had been uploaded to LSE Theses Online (LSETO)<sup>3</sup>. The decision was made to digitize theses from 2010-11 ‘backwards’ to the early 1990s. Authors were contacted and told they could opt out if they wished; only 14 chose to do so. Five take-down requests were received and complied with immediately.

The trend in downloads was upwards, and rapid. (See Figure 1.) The fast expansion of the number of theses available led to an immediate impact on downloads. The inescapable conclusion was that a higher volume of available theses attracts much more traffic. The resulting increase in the size of the repository had a temporary impact on downloads per item, which dropped briefly to an average of ten per item per month, but rapidly returned to the 2014 average of 15 downloads per item per month and at the time of writing (April 2016) is continuing to rise. Users come into the site from across the globe, with Western countries and others with large economies dominating.

‘a higher volume of available theses attracts much more traffic’

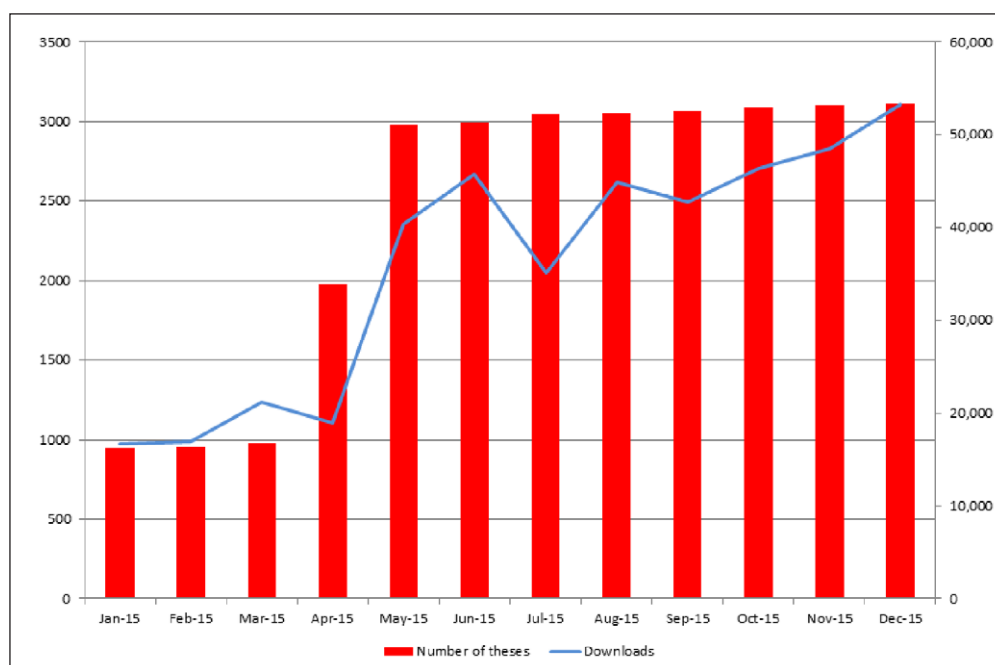


Figure 1. Relationship between number of theses in LSETO and downloads each month

## Source and objectives of visitors

Figure 2 illustrates the key sources of traffic to LSETO from January 2011 to February 2016 (on the left) and from January 2015 to February 2016 (on the right). This shows that the expansion of the repository did not substantially change access methods. The entry point for around 80% of users remained constant, with Google dominant in directing traffic. The decline in traffic share from the LSE’s own website suggests that the LSE could probably do more to promote the thesis collection on its website. Many LSE referrals come from the past PhD students’ page (e.g. for the the Department of Statistics) or the research students’ profiles (e.g. for the Department of Sociology). Others come from clicks on collection profiles on the library pages.

‘The entry point for around 80% of users remained constant, with Google dominant in directing traffic’

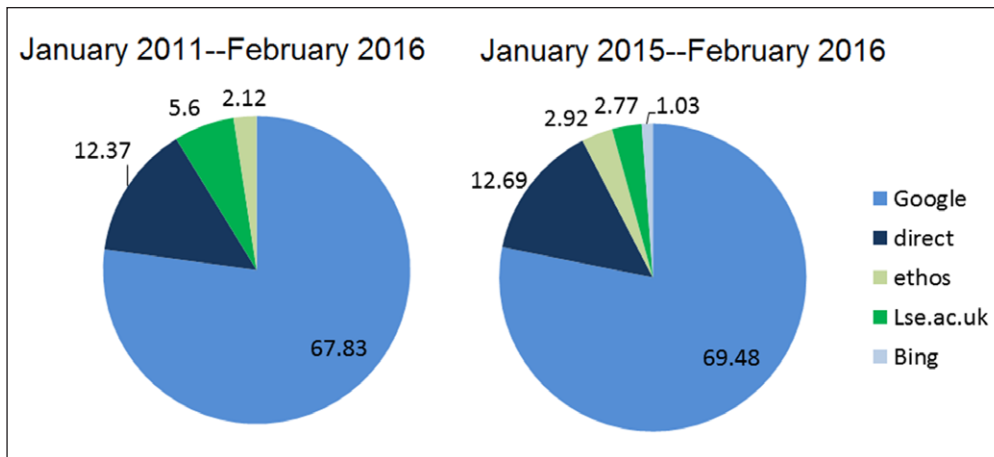


Figure 2. Largest sources of traffic to LSE Theses Online

The top searches via Google landed on the LSETO home page, which shows that many users were carrying out general searches for LSE theses, perhaps just to view an example of a thesis in their own area of research. However, many searches also led to specific theses, which both demonstrated Google’s strong indexing capabilities<sup>4</sup> and that the LSE was able to contribute scholarly research that is being sought after from its thesis collection.

The key question that the project sought to investigate was how much impact the thesis collection was having on scholarly activity. Figure 3 shows download figures and Google Scholar (GS) citations for the LSE’s top ten downloaded theses. There was no real correlation between the numbers of downloads and the citations. Even some of the older theses (more than five years old), which had had a reasonable timeframe in which to make an impact, had only achieved one or two citations.

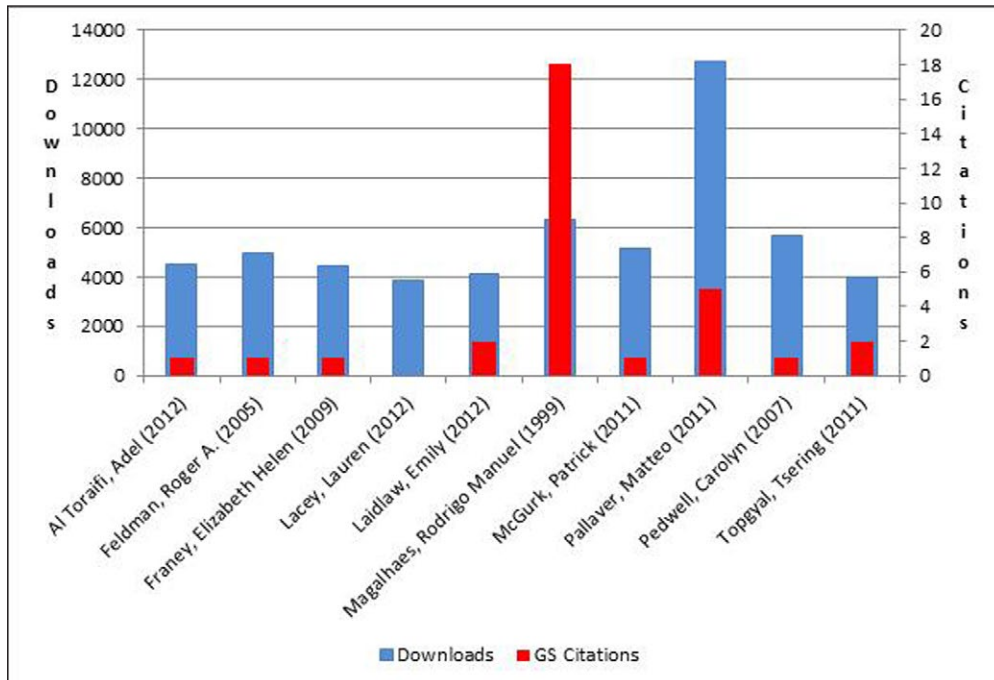


Figure 3. Top ten downloaded theses in LSE Theses Online

It is tantalizing to speculate why theses are being viewed if not because of their direct academic impact. It may be that they are on a topical subject, have a broader societal impact, or are useful for reading lists. However, further inspection of the top LSE downloaded thesis provided little further information. It appears on an international relations theory website and is cited in many foreign Masters theses, which may account for its popularity.

Figure 4 looks further at the download/citation relationship by examining the LSE theses with a strong (at least ten) citation count. The numbers of downloads for these varied considerably, and some were quite low. The cross-hatched bars show works for which there is both an original thesis and a subsequent publication with a similar title, and for which the records have been merged in GS. The cross-hatched bars stand out because they account for larger numbers of citations. This demonstrated that in order to achieve academic impact, a researcher really needs to produce a journal article/book chapter/book from the original work carried out for the thesis. However, the fact that not all have high downloads indicates that often readers do not go back to the original research output (the thesis) to investigate the work in greater depth. This therefore offers only modest proof that theses can have significant academic impact.

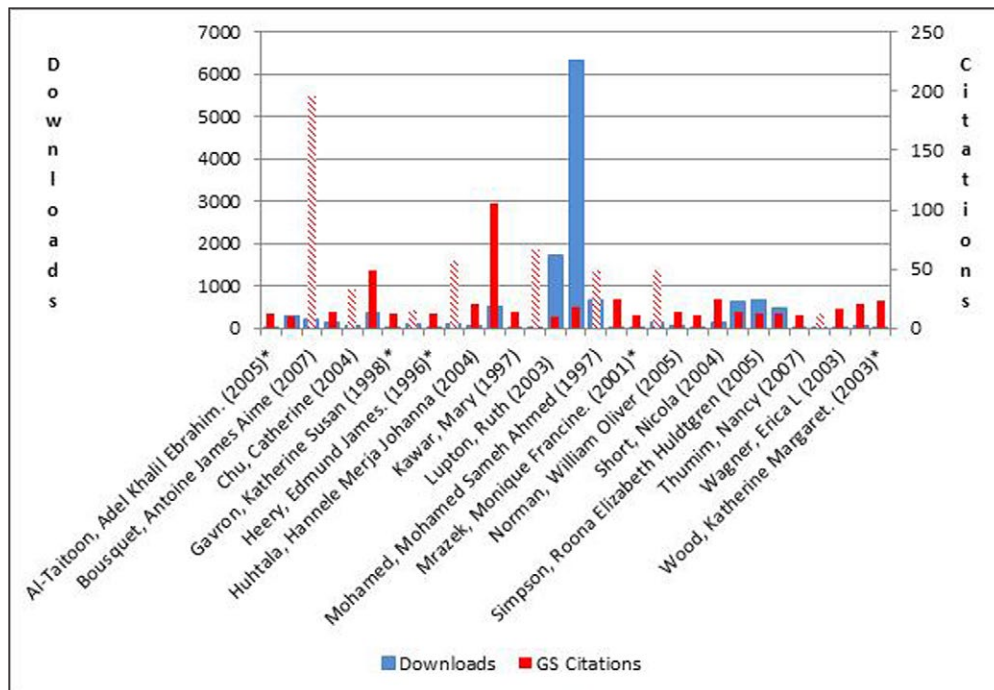


Figure 4. Sample of theses with at least ten citations according to Google Scholar (GS). The cross-hatched bars show subsequent publications which have a similar title to the thesis. \* denotes download statistics are from ETHOS as the full text is not in LSETO. NB: The Dolan thesis has a related book publication, but the records are not merged in GS as the titles are substantially different

Table 1 shows further detail on the relationship between citations and downloads. It shows the top ten downloaded theses from the ProQuest digitization project. (They had therefore been available for approximately ten months online when the research was undertaken.) The theses that had been most sought after did not necessarily have the highest citation figures. The top downloaded thesis is highly topical: it is on the Greek economy. But it was written in 1999, before Greece’s current financial problems. Perhaps it has been downloaded by researchers to get an idea of whether economists at the time were aware of nascent problems in the Greek economy. Most important for this project was that the number of times it had been downloaded made a powerful case for digitizing older theses, thus giving alumni the opportunity to reintroduce their research to contribute to relevant contemporary issues. Furthermore, if LSE theses are still important and sought-after ten, 15 or 20 years after they were submitted, this can only enhance the reputation of the institution itself.

‘theses that had been most sought after did not necessarily have the highest citation figures’

‘a powerful case for digitizing older theses’

Author	Google Scholar citations	Downloads	Published
Konsolas, Ioannis (1999) <i>The competitive advantage of nations: The case of Greece.</i>	0	1705	
Yaffe, Helen (2007), <i>Ernesto 'Che' Guevara: Socialist political economy and economic management in Cuba, 1959–1965.</i>	1	927	
Ahuja, Monika Sangeeta (1996), <i>Public interest litigation in India: A socio-legal study.</i>	1	900	
Holt, Andrew Derek (2005), <i>The role of management accounting within the development of environmental management systems.</i>	1	895	
Mohamed, Mohamed Sameh Ahmed (1997), <i>The role of the International Court of Justice as the principal judicial organ of the United Nations.</i>	49	686	Book 2003
Baxell, Richard (2002), <i>The British Battalion of the International Brigades in the Spanish Civil War, 1936–1939.</i>	1	672	
Michelutti, Lucia (2002), <i>Sons of Krishna: The politics of Yadav community formation in a north Indian town.</i>	4	520	
Joao da Costa Cabral Andresen Guimaraes, Fernando (1992), <i>The origins of the Angolan civil war. International politics and domestic political conflict 1961–1976.</i>	106	514	Book 1998
Stubb, Alexander (1999), <i>Flexible integration and the Amsterdam Treaty: negotiating differentiation in the 1996–97 IGC.</i>	12	491	
Fuller, Harcourt (2010), <i>Building a nation: Symbolic nationalism during the Kwame Nkrumah era in the Gold Coast/Ghana.</i>	1	483	

Table 1: Top ten downloaded theses from the ProQuest digitization project

This paper has already touched on methods of access to the LSE digital thesis collection. Facebook is the largest social media source that directs traffic to LSETO. This differs from the main institutional repository, LSE Research Online,<sup>5</sup> for which Twitter is the key source. One possible explanation for this is that new PhD graduates still hesitate to engage with a wider audience to promote their research, preferring to choose a platform that enables them to share their work with a smaller, closer circle.

'Facebook is the largest social media source that directs traffic to LSETO.'

Figure 5 shows two access 'spikes' that occurred. On 19 June 2015, 473 people landed on the 'browse by year' page, which suggests that they were looking for a thesis but could not find it in LSETO. It was discovered that the majority of users came from Taiwan and may have been looking for the thesis of the Presidential candidate (who was the subject of a *Time* magazine article<sup>6</sup> in which it was mentioned that she had completed her doctorate

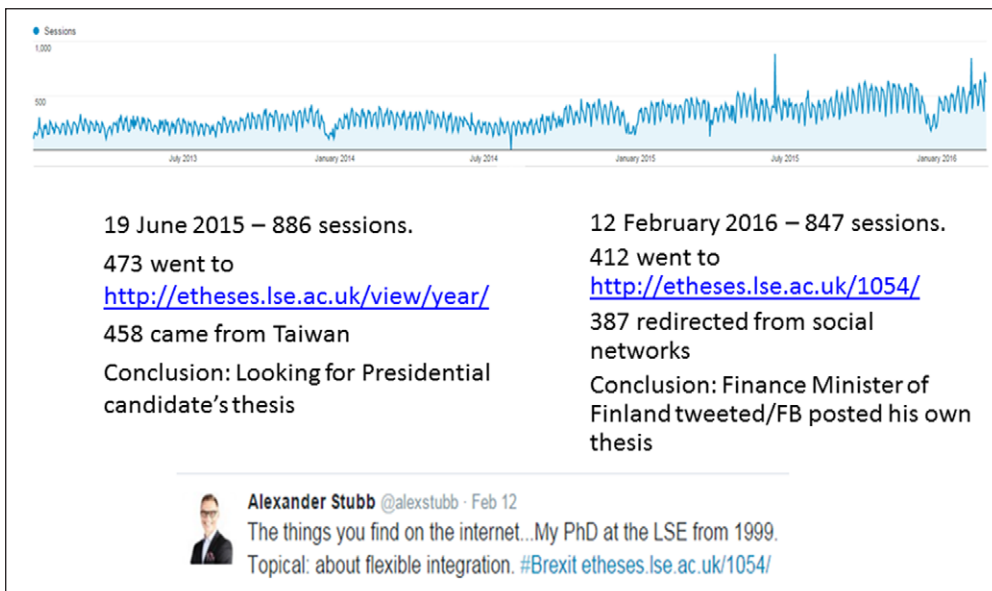


Figure 5. Spikes in traffic to LSETO

116 at LSE). The LSE Library had also received e-mail enquiries concerning the availability of the thesis. As it happened, her thesis had not been digitized; and she may or may not have been pleased if it had. The account of the focus group findings below captures the mixed response of alumni to having their theses published in this way. In the second spike, on 12 February 2016, 847 people landed on the Finance Minister of Finland's thesis after he promoted it through Facebook and Twitter. He was clearly proud of his work and wished to bring it into the 'Brexit' debate, 15 years after it was originally submitted.

## Awareness and perceptions of digital theses

Qualitative information to complement the statistics was obtained by carrying out three focus groups at the LSE, with undergraduates, postgraduates and librarians, respectively. Of the seven undergraduates who took part (none of whom was British), only four knew of the digitization project. None had been recommended to consult theses as a scholarly resource, though one had been advised to look at a digital thesis for the layout. They were enthusiastic about the opportunities that digital theses offer for accessing cutting-edge research, pointing out that they become available much more quickly than monographs or journal articles. Some also said that, since each piece of research builds on the body of work that precedes it, it is useful to have a collection of theses going back some distance in time to provide a kind of audit trail. Nevertheless, most had reservations about making their own work available as a digital thesis. Some already had ambitions to publish their research with a traditional (they actually used the word 'proper') publisher and were worried that if it appeared as a digital thesis their chances of this would be damaged. This was a concern shared also by the postgraduates and academics.

'digital theses ... become available much more quickly than monographs or journal articles'

The six postgraduate focus group participants (again, none was from the UK) were extremely motivated and quite high-powered: for example, one was working with the Bank of England, one was conducting research on far-right propaganda in France and one on food distribution and its effect on poverty in Asia. Only one knew of the digitization project and none had been told of it by their supervisors. Like the undergraduates, they believed that complying with what is now an LSE requirement to provide an electronic copy of their thesis would undermine opportunities for commercial publication. They had some other concerns, too: one thought that publishing what she described as 'academic juvenilia' might impair her reputation later on, 'when I am famous', as she put it. This group also raised concerns about copyright and permissions issues. They were unanimous in their view that the number of citations made a digital thesis of potential interest while the number of downloads did not.

The research found that the perception that publishing a thesis digitally affects future publication opportunities has almost become a *bête noire* in academic circles. The concept is so ubiquitous that it is one of the areas relating to digital theses in which a considerable amount of research has already taken place. Marisa L Ramirez, with several colleagues on each occasion, has carried out two fairly large-scale pieces of research entitled 'Do Open Access electronic theses and dissertations diminish publishing opportunities in the Sciences?'<sup>7</sup> and 'Do Open Access electronic dissertations and theses diminish publishing opportunities in the Social Sciences and Humanities?'<sup>8</sup>. Ramirez and her colleagues found that in the sciences, 'a slim majority of science journals (51.4%) reported that manuscripts derived from openly accessible electronic theses and dissertations (ETDs) are always welcome for submission, and an additional 19.4% of science journals would accept revised ETDs on a case-by-case basis', and in the social sciences and humanities (HSS), 45% of respondents considered that 'manuscripts that are revisions of openly accessible ETDs are always welcome for submission' and 27% of respondents would consider such manuscripts on a case-by-case basis. Only 12.5% of editors in the sciences and 4.5% in HSS indicated that they would under no circumstances consider such material for further publication. There is a risk, therefore, but it is nowhere near as great as most researchers imagine.

'There is a risk ... but it is nowhere near as great as most researchers imagine.'

The librarian focus group participants agreed that having the digital collection of theses was good for the reputation of both the University and the Library

117 itself – ‘Academics expect us to be there to sort out this kind of thing’ – but that better metrics were needed in order to be able to assess impact more effectively. They also pointed out that there is a lot more to building and maintaining a digital thesis collection than straightforward digitization. They, in conjunction with colleagues, have to do a lot of time-consuming training and hand-holding, especially in the form of providing support on such issues as copyright and permissions.<sup>9</sup> This obstacle was also mentioned by the librarians at Surrey, who said that although Surrey authors have been instructed about third-party copyright for many years, the Surrey project highlighted that the rules had often been disregarded. Managing expectations was also a problem: ‘We have 3,500 digitized theses. We have that number again and more that haven’t been digitized. This means that we can’t manage expectations.’

Of the four academics who took part in the semi-structured telephone calls, two knew of the digitization project. Two were in favour of digital theses, while the other two had reservations, again connected with copyright and publication issues. Three had consulted digital theses and all could see where their value lay: helping promote cutting-edge research and putting new research in a historical context were particularly mentioned. Like the postgraduates, they were interested only in citations, not in downloads.

‘interested only in citations, not in downloads’

Figure 6 shows a comparison between the LSE’s usage statistics and Surrey’s. The pattern is very similar, showing that use of digital theses mirrors the academic cycle (quieter in the summer). They also show that these two very different institutions attracted more or less the same level of traffic per item over time. (However, Surrey’s top ten downloads attracted a much higher volume of hits than the LSE’s).

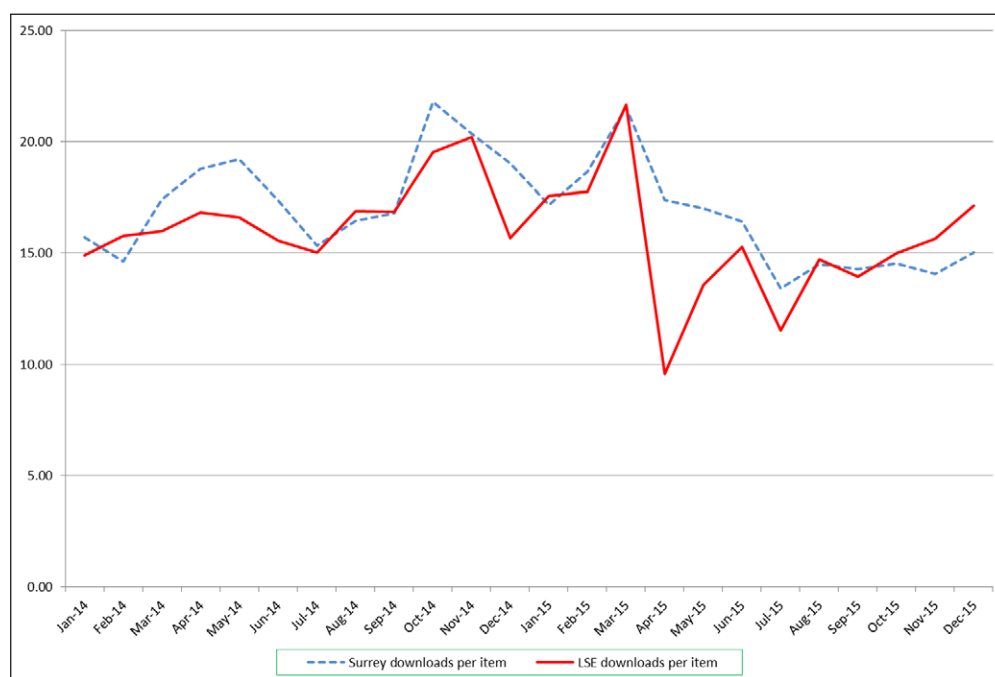


Figure 6. Downloads per item for LSE and University of Surrey

## Conclusions

Digitization of theses has brought many more users to LSETO. It is difficult to assess with any accuracy the impact this has had on scholarly research: there seems to be no direct relationship between downloads and citations. If researchers wish to make an impact on the body of scholarly resources, subsequent publication in a book or journal seems to be of more consequence (and other research suggests that making the thesis available digitally has a relatively minor effect on opportunities to do this). However, impact on the micro level should still be considered important. The LSE’s alumni deserve the chance that a digital thesis collection offers of bringing their work to the debate again when it may have languished in unjustified obscurity because of more limited opportunities to promote it when they first submitted their theses.

118 The project demonstrated that more discussions need to take place between librarians and academic departments to enable better promotion of theses. The LSE has used the evidence that very few authors opted out of the ProQuest digitization project and few take-down requests were received to push for a change in its policy with EThOS (the e-theses online service). It has adopted the policy of many other UK universities, including Surrey, and now no longer chases author permissions, but instead views digitization as merely a format change which requires no permission and relies on a take-down policy. This will allow for a greater inclusion of older theses and generate repository growth.

'merely a format change which requires no permission and relies on a take-down policy'

The LSE and Surrey, with ProQuest as a partner, hope to take this research forward. ProQuest understands that better data, as well as better promotion of digitization of theses projects are important in order to find out more about how digital theses are being used as a scholarly resource, and continues to improve its PQDT statistics. The relationship between publishing a digital thesis and opportunities for converting it to a traditionally-published monograph need to be better understood and Surrey, in conjunction with the UK Publishers Association, has begun work on this. Universities and university libraries need more help with explaining permissions, copyright and intellectual property rights, embargoes and other author and publication issues. As well as being important in its own right, to enable maximization of use of theses as a scholarly resource, it is believed that future work will contribute directly to the impact of the open access movement.<sup>10</sup>

#### Acknowledgements

The authors would like to thank ProQuest and Fiona Greig, Head of Strategy and E-resources at the University of Surrey Library, for their support of this project.

#### Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'Abbreviations and Acronyms' link at the top of the page it directs you to: <http://www.uksg.org/publications#aa>

#### Competing Interests

The authors have declared no competing interests.

#### References

- 1 QS Top Universities: <http://www.topuniversities.com/university-rankings/world-university-rankings/2015#sorting=rank+region+=+country+=+faculty+=+stars=false+search> (accessed 3 June 2016).
- 2 There is some reference to digital thesis collections in Rosenberg Z, Citation analysis of MA theses and PhD dissertations in Sociology and Anthropology, *The Journal of Academic Librarianship*, 2015, 41 (680–688); DOI: <http://dx.doi.org/10.1016/j.acalib.2015.05.010> (accessed 9 June 2016), but the paper discusses the use of theses and dissertations within the wider context of all the library's digital resources.
- 3 LSE Theses Online: <http://etheses.lse.ac.uk> (accessed 9 June 2016).
- 4 Discoverability of electronic theses and dissertations is the subject of McCutcheon S, Kreyche M and Maurer M B, Morphing metadata: maximizing access to electronic theses and dissertations, *Library Hi Tech*, 2008, 26 (1) 41–57; DOI: <http://dx.doi.org/10.1108/07378830810857799> (accessed 9 June 2016).
- 5 LSE Research Online: <http://eprints.lse.ac.uk/> (accessed 13 June 2016).
- 6 'The Next President Of Taiwan', *Time*, 18 June 2015. Available from: <http://time.com/3926330/the-next-president-of-taiwan/> (accessed 3 June 2016).
- 7 See Ramirez M L, McMillan G, Dalton J T, Halon A, Smith H S and Kern C, Do Open Access electronic theses and dissertations diminish publishing opportunities in the Sciences?, *College and Research Libraries*, Early view: <http://crl.acrl.org/content/early/2013/10/29/crl13-524.short>.
- 8 See Ramirez M L, Dalton JT, McMillan G, Read M and Seamans N, Do Open Access Electronic Dissertations and Theses Diminish Publishing Opportunities in the Social Sciences and Humanities? Findings from a 2011 Survey of Academic Publishers, *College and Research Libraries*, crl 356–380; DOI: <http://dx.doi.org/10.5860/crl-356> (accessed 9 June 2016).
- 9 For a discussion on how much work has to go into building and maintaining a digital thesis collection, see Han Y, ETD: total cost of ownership – collecting, archiving and providing access, *Library Management*, 2014, 35(4/5), 250–258; DOI: <http://dx.doi.org/10.1108/LM-08-2013-0084> (accessed 9 June 2016).
- 10 One aspect of the relationship of OA theses and dissertations to the OA movement in general has been considered by Schopfel J, Chaudiron S and Jacquemin B, Open access to research data in electronic theses and dissertations: an overview, *Library Hi Tech*, 2014, 32(4), 612–617; DOI: <http://dx.doi.org/10.1108/LHT-06-2014-0058> (accessed 9 June 2016).



**Article copyright: © 2016 Linda Bennett and Dimity Flanagan. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.**



Corresponding author: Linda Bennett

Independent Researcher: Gold Leaf, GB

E-mail: [lbennett@goldleaf.co.uk](mailto:lbennett@goldleaf.co.uk)

Co-author: Dimity Flanagan

ORCID iD: <http://orcid.org/0000-0002-6784-0325>

To cite this article:

Bennett, L and Flanagan, D, Measuring the impact of digitized theses: a case study from the London School of Economics, *Insights*, 2016, 29(2), 111–119; DOI: <http://dx.doi.org/10.1629/uksg.300>

Published by UKSG in association with Ubiquity Press on 05 July 2016